Three-dimensional forward and backward numerical modeling of mantle plume evolution: Effects of thermal diffusion

Alik Ismail-Zadeh, Gerald Schubert, Igor Tsepelev, Alexander Korotkii

Three-dimensional forward and backward numerical modeling of mantle plume evolution: Effects of thermal diffusion

Alik Ismail-Zadeh,[1,2,3] Gerald Schubert,[4] Igor Tsepelev,[5] and Alexander Korotkii[5]

Received 14 April 2005; revised 6 February 2006; accepted 22 February 2006; published 8 June 2006.

[1] We investigate the effects of thermal diffusion on the evolution of mantle plumes by means of three-dimensional numerical modeling forward and backward in time. Mantle plumes are fed by a hot, low-viscous material from the thermal boundary layer. The material of the plumes is mainly advected toward the Earth's surface with some effects of thermal diffusion. However, the feeding can become weaker with time, and then thermal diffusion can take over and control the evolution of the plumes. Numerical experiments forward in time show that a week feeding of mantle plumes by the hot material from the boundary layer results in the diffusive disappearance of plume tails first and plume heads later. This is the most likely explanation for the seismically detected low-velocity mantle structures (mantle plumes) with prominent heads and almost invisible tails at midmantle depths. We develop restoration models (backward in time) to recover strong features of mantle plumes in the geological past after they have dissipated due to thermal diffusion and analyze effects of thermal diffusion and temperature-dependent viscosity on the reconstruction of the mantle plumes. We investigate the impact of thermal diffusion on the performance of our restoration (variational data assimilation) algorithm. For a given range of Rayleigh number *Ra* and two values of the viscosity ratio *r* (between the upper and lower boundaries of the model domain) we show that (1) the residuals between the temperature predicted by the forward model and that reconstructed by the backward modeling become larger and (2) the restoration process becomes poorer as *Ra* decreases and *r* increases. We assimilate temperature obtained from high-resolution seismic tomography data for the southeastern Carpathians and show that present diffused mantle structures can be restored to their prominent state in the Miocene times. We discuss the problems of smoothness of model input and output data, errors associated with the modeling, and some other challenges in the data assimilation for thermoconvective flow in the mantle.

Citation: Ismail-Zadeh, A., G. Schubert, I. Tsepelev, and A. Korotkii (2006), Three-dimensional forward and backward numerical modeling of mantle plume evolution: Effects of thermal diffusion, *J. Geophys. Res.*, *111*, B06401, doi:10.1029/2005JB003782.

1. Introduction

[2] Mantle plumes are among the most spectacular features of mass and heat transport from the mantle to the Earth's surface. Thermal plumes in the mantle plausibly originate near either the core-mantle boundary or the upper mantle–lower mantle transition due to instabilities in the hot thermal boundary layers that could exist at these locations. Although some mantle plumes appear to last for more than 150 Myr, they are nonetheless transient features: no tracks older than the Mesozoic are well established [e.g., *Condie*, 2001; *Jellinek and Manga*, 2002]. Direct observational evidence of mantle plumes comes from seismic tomography, which provides constraints on temperature and composition of present mantle structures [e.g., *Ritsema et al.*, 1999; *Montelli et al.*, 2004]. Our understanding of mantle plume dynamics comes from numerical (*Schubert et al.* [2001] provide an overview) and laboratory [e.g., *Davaille*, 1999] experiments.

[3] Numerical models of mantle plume evolution have been mainly carried out forward in time, i.e., from the onset of plumes to late stages of maturity. The main drawback of these models is that the initial conditions (conditions in the geological past) for the models are unknown. However, temperature and flow at the time of plume onset can be inferred from the present mantle temperature and flow using data assimilation based on combined forward and backward numerical modeling of plume evolution. The main motivation for the data assimilation comes from the rapid progress

[1]International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, Moscow, Russia.
[2]Geophysikalishes Institut, Universität Karlsruhe, Karlsruhe, Germany.
[3]Institut de Physique du Globe de Paris, Paris, France.
[4]Department of Earth and Space Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, USA.
[5]Institute of Mathematics and Mechanics, Ural Branch, Russian Academy of Sciences, Yekaterinburg, Russia.

0148-0227/06/2005JB003782$09.00

made by seismic tomographers in imaging deep Earth structure. Restoration of seismically imaged structures backward in time could provide an important way to test a range of geodynamics hypotheses.

[4] Data assimilation is defined as the incorporation of present (observations) and past (initial conditions) data in an explicit dynamical model to provide time continuity and coupling among the physical fields. The basic principle of data assimilation is to consider the initial condition as a control variable and to optimize the initial condition in order to minimize the discrepancy between the observations and the solution of the model. Data related to a thermoconvective mantle flow can be assimilated by using sequential filtering, variational technique, and some others methods [e.g., *Ghil and Malanotte-Rizzoli*, 1991; *Wunsch*, 1996; *Talagrand*, 1997]. In sequential filtering a numerical model is computed forward in time for the interval for which observations have been made, updating the model each time where observations are available. *Bunge et al.* [1998, 2002] used this approach to compute mantle circulation models. Despite sequential data assimilation well adapted to mantle circulation studies, each individual observation influences the model state at later times. Information propagates from the geological past into the future, although our knowledge of the Earth's mantle at earlier times is much poor than that at present.

[5] The use of variational data assimilation in solid Earth dynamics (to estimate initial mantle temperature and flow in the geological past) has been put forward by *Bunge et al.* [2003] and *Ismail-Zadeh et al.* [2003a, 2003b]. This idea is based on a variational technique applied to solve the coupled heat, momentum and continuity equations in order to find the model representation that is most consistent with the observations. That best estimate can then be used to analyze geodynamic processes or initialize a model setup more accurately. *Ismail-Zadeh et al.* [2004] presented a data assimilation algorithm for numerical restoration of a three-dimensional model of present prominent mantle plumes to past stages and showed a high-accuracy in recovering the initial configurations of these plumes. The two major objectives of this study are (1) to estimate effects of thermal diffusion and temperature-dependent viscosity on the evolution of mantle plumes and (2) to recover the structure of mantle plumes prominent in the past from that of present plumes weakened by thermal diffusion.

[6] Conduction and convection are two major mechanisms for the transfer of heat. Conductive heat transfer in the mantle is a diffusion process occurring due to collisions of molecules, which transmit their kinetic energies to other molecules. Convective heat transfer is associated with the mantle motion due to buoyancy and plays a dominant part in the general transport of heat from the deep interior of the Earth to the surface. The thermal conductivity of mantle material depends on pressure and temperature. A model for thermal conductivity in the sublithospheric mantle, based on the experimental study (photon lifetimes obtained from infrared reflectivity) by *Hofmeister* [1999], shows that the thermal conductivity increases with depth from about 2 to 7 W m^{-1} K^{-1}. In addition to transport by conduction and convection, a hot material produces blackbody radiation, and heat is diffused if the light emitted by one particle is partially scattered or absorbed by high-frequency transitions in neighboring molecules. *Badro et al.* [2004] showed experimentally a substantial increase in radiative thermal conductivity in the lower mantle. The change in the radiative conductivity of lower mantle minerals will influence the lower mantle dynamics and plume evolution, because the increase in thermal conductivity results in a decrease of the Rayleigh number and hence in an increase of thermal diffusion.

[7] We briefly describe the model setup and numerical method in section 2 and the variational data assimilation approach to the reconstruction of mantle plumes in section 3 (details of this approach are given by *Ismail-Zadeh et al.* [2004]). We present three-dimensional forward numerical models of mantle plume weakening due to thermal diffusion and analyze the influence of thermal diffusion and viscosity ratio on the evolution of mantle plumes in section 4. These diffused plume structures are then restored to their prominent state in the past, and we analyze the effects of thermal diffusion and viscosity on the reconstruction of mantle plumes in section 5. The efficiency of the data assimilation technique is illustrated in section 6 in terms of the number of iterations required to obtain the target temperature and flow velocity in the past. In section 7.1 we discuss how the numerical results on fading mantle plumes can explain the recent seismic tomography observations of low-velocity anomalies extending down to midmantle depths. We show in section 7.2 the applicability of the numerical reconstruction method (data assimilation approach) to "real" (that is, imaged by seismic tomography) mantle structures, and present conclusions in section 8.

2. Model Problem and Numerical Approach

[8] We study the problem of mantle plume evolution in the three-dimensional model domain $\Omega = [0, x_1 = 3h] \times [0, x_2 = 3h] \times [0, x_3 = h]$, where $\mathbf{x} = (x_1, x_2, x_3)$ are the Cartesian coordinates and h is the depth of the domain. We assume that the mantle behaves as a Newtonian incompressible fluid with a temperature-dependent viscosity and infinite Prandtl number. Rising mantle plumes are modeled as hot fluid jets ascending into the relatively cold ambient fluid heated from below. The mantle flow is described by heat, motion, and continuity equations [*Chandrasekhar*, 1961]. To simplify the governing equations, we make the Boussinesq approximation [*Boussinesq*, 1903] keeping the density constant everywhere except for buoyancy term in the equation of motion. We note that a variable (temperature-dependent) density [*Ismail-Zadeh et al.*, 2003a] and an internal heating [*Bunge et al.*, 2003] can be also used in the forward and backward modeling of thermoconvective mantle circulation. In the Boussinesq approximation the dimensionless equations take the form

$$\partial T/\partial t + \mathbf{u} \cdot \nabla T = \nabla^2 T, \quad t \in (0, \vartheta), \quad \mathbf{x} \in \Omega, \tag{1}$$

$$\nabla P = \mathrm{div}[\eta(T)\mathbf{E}] + RaT\mathbf{e}, \quad \mathbf{E} = \{\partial u_i/\partial x_j + \partial u_j/\partial x_i\}, \quad \mathbf{e} = (0, 0, 1), \tag{2}$$

$$\mathrm{div}\mathbf{u} = 0, \quad t \in (0, \vartheta), \quad \mathbf{x} \in \Omega. \tag{3}$$

Here T, t, $\mathbf{u} = (u_1, u_2, u_3)$, P, and η are dimensionless temperature, time, velocity, pressure, and viscosity, respectively. The Rayleigh number is defined as $Ra = \alpha g \rho_{\mathrm{ref}} \Delta T h^3 \eta_{\mathrm{ref}}^{-1} \kappa^{-1}$, where α is the thermal expansivity, g is the acceleration due to gravity, ρ_{ref} and η_{ref} are the reference typical density and viscosity, respectively; ΔT is the temperature contrast between the lower and upper boundaries of the model domain; and κ is the thermal diffusivity. In equations (1)–(3), length, temperature, and time are normalized by h, ΔT, and $h^2\kappa^{-1}$, respectively.

[9] At the boundary of the model domain we set the impenetrability condition with perfect slip conditions: $\partial \mathbf{u}_\tau / \partial \mathbf{n} = 0$, $\mathbf{u} \cdot \mathbf{n} = 0$, where $\mathbf{n}$ is the outward unit normal vector at a point on the model boundary, and $\mathbf{u}_\tau$ is the projection of the velocity vector onto the tangent plane at the same point on the model boundary. We assume zero heat flux through the vertical boundaries of the box. The upper and lower boundaries are isothermal surfaces, and we set $T = 0$ and $T = 1$ at these boundaries, respectively.

[10] Equations (1)–(3) together with the boundary conditions describe a thermoconvective mantle flow. To solve the problem forward or backward in time we assume the temperature to be known at the time of plume onset ($t = 0$) or at the present time ($t = \vartheta$).

[11] Temperature in the heat equation (1) is approximated by finite differences and determined by the semi-Lagrangian method, which allows for relatively large time steps, high accuracy, and low numerical diffusion [*McDonald*, 1984]. A numerical solution to the Stokes equations (2) is based on the introduction of a two-component vector velocity potential and on the application of the Eulerian finite element method with a tricubic-spline basis for computing the potential [*Ismail-Zadeh et al.*, 2001]. Such a procedure results in a set of linear algebraic equations with a symmetric positive-definite banded matrix. We solve the set of equations by the conjugate gradient method [*Fletcher and Reeves*, 1964]. The numerical algorithm was designed to be implemented on parallel computers. The reader is referred to *Ismail-Zadeh et al.* [2001, 2004] for more detail.

3. Variational Data Assimilation

[12] Data assimilation techniques has been pioneered by meteorologists and used very successfully to improve operational weather forecasts [e.g., *Kalnay*, 2003]. Data assimilation has also been widely used in oceanography [e.g., *Bennett*, 1992] and in hydrological studies [e.g., *McLaughlin*, 2002]. However, the application of the method to problems of mantle dynamics is still in its infancy.

[13] The variational data assimilation is based on a search of the best fit between the forecast model state and the observations by minimizing an objective functional (a normalized residual between the target model and observed variables) over space and time. To minimize the objective functional over time, an assimilation time interval is defined and an adjoint model is typically used to find the derivatives of the objective functional with respect to the model states. The variational data assimilation is well suited for smoothing problems (we discuss the problem of smoothness of the initial data and solution in Appendix A).

[14] The method for variational data assimilation can be formulated with a weak constraint where errors in the model formulation are taken into account as control parameters (generalized inverse) [*Bunge et al.*, 2003] or with a strong constraint where the model is assumed to be perfect except for the errors associated with the initial conditions [*Bunge et al.*, 2003; *Ismail-Zadeh et al.*, 2003a]. There are several sources of errors in forward and backward modeling of thermoconvective mantle flow, which we discuss in Appendix B. The generalized inverse of mantle convection considers model errors, data misfit and the misfit of parameters as control variables. Unfortunately the generalized inverse presents a tremendous computational challenge and is difficult to solve in practice. Hence *Bunge et al.* [2003] considered a simplified generalized inverse imposing a strong constraint on errors (ignoring all errors except for the initial condition errors). Therefore the strong constraint makes the problem computationally tractable.

[15] We consider the following objective functional $J(\varphi) = \|T(\vartheta, \cdot\ ; \varphi) - \chi(\cdot)\|^2$, where parallels denote the norm in the space L_2 (Ω) (the Hilbert space with the norm defined as $\|y\| = [\int_\Omega y^2(\mathbf{x}) d\mathbf{x}]^{1/2}$). Since in what follows the dependence of solutions of the thermal boundary value problems on initial data is important, we introduce these data explicitly into the mathematical representation of temperature. Here $T(\vartheta, \cdot\ ; \varphi)$ is the solution of the thermal boundary value problem (1) at the final time ϑ, which corresponds to some (unknown as yet) initial temperature distribution $\varphi(\mathbf{x})$; $\chi(\mathbf{x}) = T(\vartheta, \mathbf{x}; T_0)$ is the known temperature distribution at the final time, which corresponds to the initial temperature $T_0(\cdot)$. The functional has its unique global minimum at value $\varphi \equiv T_0$ and $J(T_0) \equiv 0$, $\nabla J(T_0) \equiv 0$. To find the minimum of the functional we employ the gradient method ($k = 0, \ldots, j, \ldots$):

$$\varphi_{k+1} = \varphi_k - \beta_k \nabla J(\varphi_k), \tag{4}$$

$$\beta_k = \min\{1/(k+1), J(\varphi_k)/\|\nabla J(\varphi_k)\|\}, \quad \varphi_0 = T_*, \tag{5}$$

where T_* is an initial temperature guess. The minimization method belongs to a class of limited-memory quasi-Newton methods [*Zou et al.*, 1993], where approximations to the inverse Hessian matrices are chosen to be the identity matrix. The gradient of the objective functional $\nabla J(\varphi_k)$ decreases steadily with the number of iterations, and it provides the convergence of the method. Meanwhile the absolute value of β_k increases with the number of iterations, and it can result in instability of the iteration process [*Samarskii and Vabischevich*, 2004]. To avoid the instability, we use equation (5) to minimize the parameter β_k.

[16] The minimization algorithm requires the calculation of the gradient of the objective functional, ∇J. This can be done through the use of the adjoint problem for the model equations (1)–(3) with the relevant boundary and initial conditions. In the case of the heat problem, the adjoint problem can be represented in the following form:

$$\begin{aligned} &\partial Z/\partial \tau - \mathbf{u} \cdot \nabla Z = \nabla^2 Z, \qquad \tau = \vartheta - t \in (-\vartheta, 0), \\ &Z(0, \mathbf{x}) = 2(T(\vartheta, \mathbf{x}; \varphi) - \chi(\mathbf{x})), \qquad \mathbf{x} \in \Omega, \end{aligned} \tag{6}$$

with uniform boundary conditions. The solution $Z(\vartheta, \cdot)$ to this adjoint problem is the gradient of the objective functional, and

the gradient is derived by using the Fréchet derivative of the functional [see *Ismail-Zadeh et al.*, 2004, Appendix B]. The correctness of the solution has been verified by the gradient accuracy test [*Navon et al.*, 1992, equation 2.20].

[17] We define a uniform partition of the time axis at points $t_n = \vartheta - n\delta t$, where δt is the time step, and n successively takes integer values from 0 to some natural number $m = \vartheta/\delta t$. At each subinterval of time $[t_{n+1}, t_n]$, the search of the temperature T and flow velocity $\mathbf{u}$ at $t = t_{n+1}$ consists of the following basic steps.

[18] 1. Given the temperature $T = T(t_n, \mathbf{x})$ at $t = t_n$ we solve a set of linear algebraic equations derived from equations (2) and (3) with the appropriate boundary conditions in order to determine the velocity $\mathbf{u}$.

[19] 2. The "advective" temperature $T_{adv} = T_{adv}(t_{n+1}, \mathbf{x})$ is determined by solving the advection heat equation backward in time, neglecting the diffusion term in equation (1). This can be done by replacing positive time steps by negative ones [see *Ismail-Zadeh et al.*, 2003b].

[20] Given the temperature $T = T_{adv}$ at $t = t_{n+1}$ steps 1 and 2 are then repeated to find the velocity $\mathbf{u}_{adv} = \mathbf{u}(t_{n+1}, \mathbf{x}; T_{adv})$.

[21] 3. The heat equation (1) is solved with appropriate boundary conditions and initial condition $\varphi_0(\mathbf{x}) = T_{adv}(t_{n+1}, \mathbf{x})$ forward in time using velocity $\mathbf{u}_{adv}$ in order to find $T(t_n, \mathbf{x}; \varphi_0)$.

[22] 4. The adjoint equation (6) is then solved backward in time with appropriate boundary conditions and initial condition $T(t_n, \mathbf{x}) = T(t_n, \mathbf{x}; \varphi_0)$ using velocity $\mathbf{u}$ in order to determine $\nabla J(\varphi_0)$.

[23] 5. The coefficient β_0 is determined from equation (5), and the temperature is updated (i.e., φ_1 is determined) from equation (4).

[24] Steps 3 to 5 are repeated for φ_j and β_j ($j = 1, 2, 3, \ldots$) until $\delta\varphi_j = J(\varphi_j) + \|\nabla J(\varphi_j)\|^2 < \varepsilon$, where ε is a small constant. Temperature φ_j is then considered to be the approximation to the target value of the initial temperature $T(t_{n+1}, \mathbf{x})$. Finally, step 1 is used to determine the flow velocity $\mathbf{u}(t_{n+1}, \mathbf{x}; T(t_{n+1}, \mathbf{x}))$.

[25] Step 2 introduces a preconditioner to accelerate the convergence of temperature iterations in steps 3 to 5 at high Rayleigh number. At low *Ra*, step 2 is omitted and $\mathbf{u}_{\mathrm{adv}}$ is replaced by $\mathbf{u}$.

4. Forward Modeling of Mantle Plume Diffusion

[26] Mantle plumes evolve in three distinguishing stages: (1) immature, i.e., an origin and initial rise of the plumes; (2) mature, i.e., plume-lithosphere interaction, gravity spreading of plume head and development of overhangs beneath the bottom of the lithosphere, and partial melting of the plume material [e.g., *Ribe and Christensen*, 1994; *Moore et al.*, 1998]; and (3) overmature, i.e., slowing down of the plume rise and fading of the mantle plumes due to thermal diffusion [*Davaille and Vatteville*, 2005]. The ascent and evolution of mantle plumes depend on the properties of the source region (that is, the thermal boundary layer) and the viscosity and thermal diffusivity of the ambient mantle. The properties of the source region determine temperature and viscosity of the mantle plumes. Structure, flow rate, and heat flux of the plumes are controlled by the properties of the mantle through which the plumes rise. While properties of the lower mantle (e.g., viscosity, thermal conductivity) are relatively constant during about 150 Myr lifetime of most plumes, source region properties can vary substantially with time as the thermal basal boundary layer feeding the plume is depleted of hot material. Complete local depletion of this boundary layer cuts the plume off from its source. It is the subsequent evolution of the plume that interests us here.

[27] We study only the late stage of the mantle plume evolution associated with the fading of the plume due to thermal diffusion and model the evolution of mantle plumes deprived of source material through numerical experiments of three-dimensional thermal convection in a bottom heated box. The mantle behaves as a Newtonian fluid on geological timescales, and a dimensionless temperature-dependent viscosity law [*Busse et al.*, 1993] given by

$$\eta(T) = \exp\left(\frac{M}{T+G} - \frac{M}{0.5+G}\right)$$

is used in the modeling, where $M = [225/\ln(r)] - 0.25\ln(r)$, $G = 15/\ln(r) - 0.5$ and r is the viscosity ratio between the upper and lower boundaries of the model domain. We model the evolution of mantle plumes for two viscosity profiles: $r = 20$ and $r = 200$. The temperature-dependent viscosity profile has its minimum at the core-mantle boundary. A more realistic viscosity profile [e.g., *Forte and Mitrovica*, 2001] will influence the evolution of mantle plumes, though it will not influence the restoration of the plumes. The model domain is divided into $37 \times 37 \times 29$ rectangular finite elements to approximate the vector velocity potential by tricubic splines, and a uniform grid $112 \times 112 \times 88$ is employed for approximation of temperature, velocity, and viscosity.

[28] Initially, we model the evolution of mature mantle plumes to obtain initial temperature data for models of mantle plume diffusion. With $\alpha = 3 \times 10^{-5}$ K^{-1}, $\rho_{ref} = 4000$ kg m^{-3}, $\Delta T = 3000$ K, $h = 2800$ km, $\eta_{ref} = 8 \times 10^{22}$ Pa s, and $\kappa = 10^{-6}$ m^{-2} s^{-1}, the initial Rayleigh number is $\mathrm{Ra} = 9.5 \times 10^5$. While plumes evolve in the convecting heterogeneous mantle, at the initial time we assume that the plumes develop in a laterally homogeneous temperature field and hence consider that the mantle temperature in the model increases linearly with depth.

[29] Mantle plumes are generated by random temperature perturbations at the top of the thermal source layer associated with the core-mantle boundary (Figure 1a). The mantle material in the basal source layer flows horizontally toward the plumes. The reduced viscosity in this basal layer promotes the flow of the material to the plumes. Vertical upwelling of hot mantle material is concentrated in low-viscosity conduits near the centerlines of the emerging plumes (Figures 1b and 1c). The plumes move upward through the model domain, gradually forming structures with well-developed heads and tails. The plumes diminish in size with time (Figure 1d), and the plume tails disappear before the plume heads (Figures 1e and 1f). We note that the figures present a hot isothermal surface of the plumes. If colder isotherms are considered, the disappearance of the isotherms will occur later. However, anyhow, hot or cold isotherms are plotted, plume tails will vanish before their heads. Results of recent laboratory experiments [*Davaille and Vatteville*, 2005] support strongly our numerical find-

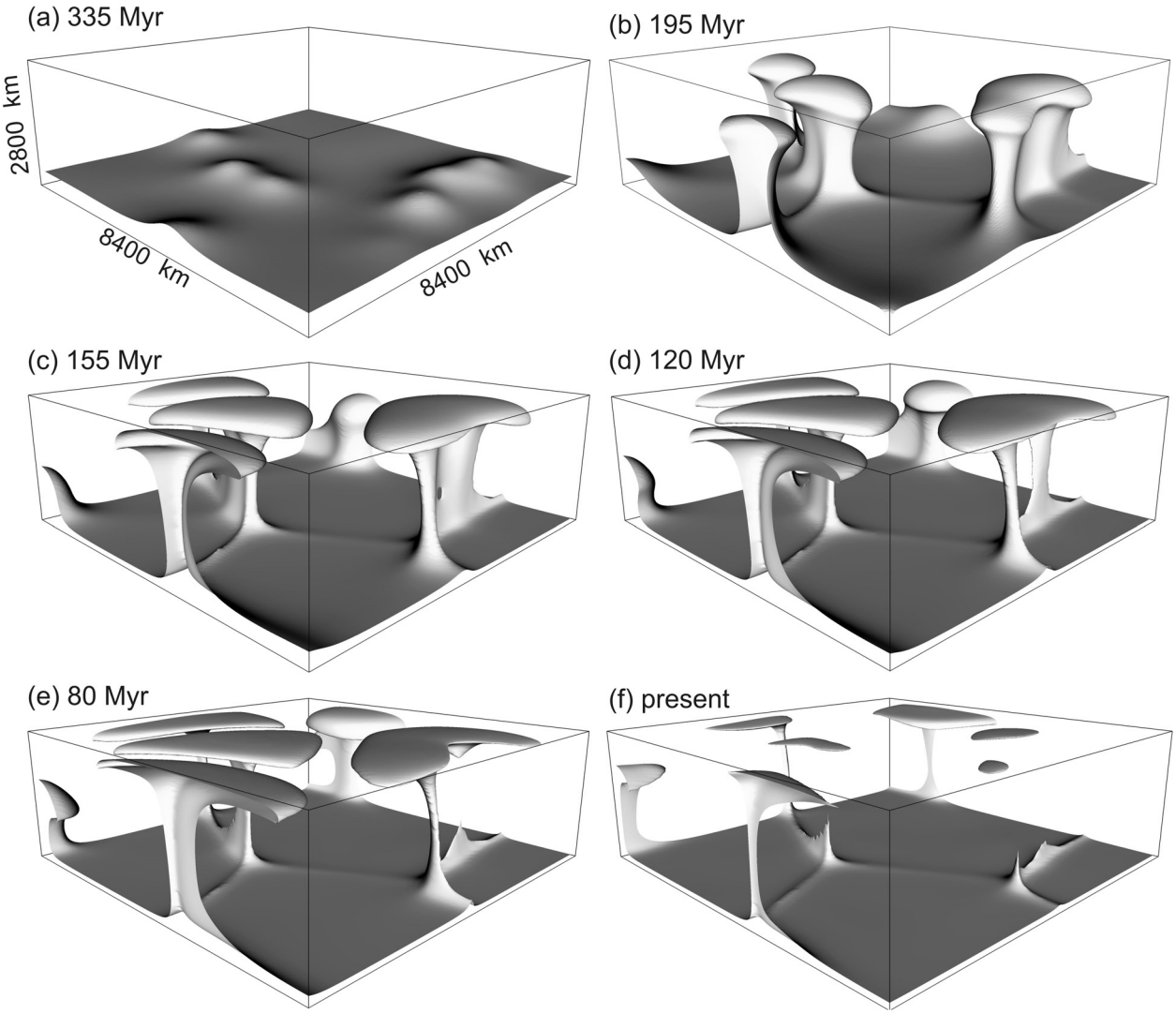

Figure 1. Mantle plumes in the forward modeling at successive diffusion times: from (a) 335 Myr ago to (f) the "present" state of the plumes. The plumes are represented here and in Figures 2 and 3 by isothermal surfaces at 3000 K.

ings that plumes start disappearing from bottom up and fade away by thermal diffusion.

[30] At different stages in the plume decay one sees quite isolated plume heads, plume heads with short tails, and plumes with nearly pinched off tails. Different amounts of time are required for different mantle plumes to vanish into the ambient mantle, the required time depending on the geometry of the plume tails. Temperature loss is greater for sheet-like tails than for cylindrical tails. The tails of the cylindrical plumes (e.g., Figure 1c, in the left part of the model domain) are still detectable after about 155 Myr. However, at this time the sheet-like tail of the large plume in the right part of the model domain (Figure 1c) is already invisible and only its head is preserved in the uppermost mantle (Figure 1f). Two-dimensional numerical experiments of steady state convection [*Leitch et al.*, 1996] reveal a significant change in the centerline temperature of sheet-like plume tails compared to the cylindrical plume tail due to heat conduction in the horizontal direction.

5. Recovering Prominent Mantle Plumes From Their Weakened Present Stage

[31] We use the numerical approach described in section 3 to reconstruct the prominent state of the plumes (Figure 1d) in the past from their "present" weak state (Figure 1f). Figure 2 illustrates the reconstructed states of the plumes (Figures 2e–2g) and the temperature residuals δT (Figures 2h–2j) between the temperature $T(\mathbf{x})$ predicted by the forward model and the temperature $\widetilde{T}(\mathbf{x})$ reconstructed to the same age:

$$\delta T(x_1, x_2) = \left[\int_0^h \left(T(x_1, x_2, x_3) - \widetilde{T}(x_1, x_2, x_3) \right)^2 dx_3 \right]^{1/2}.$$

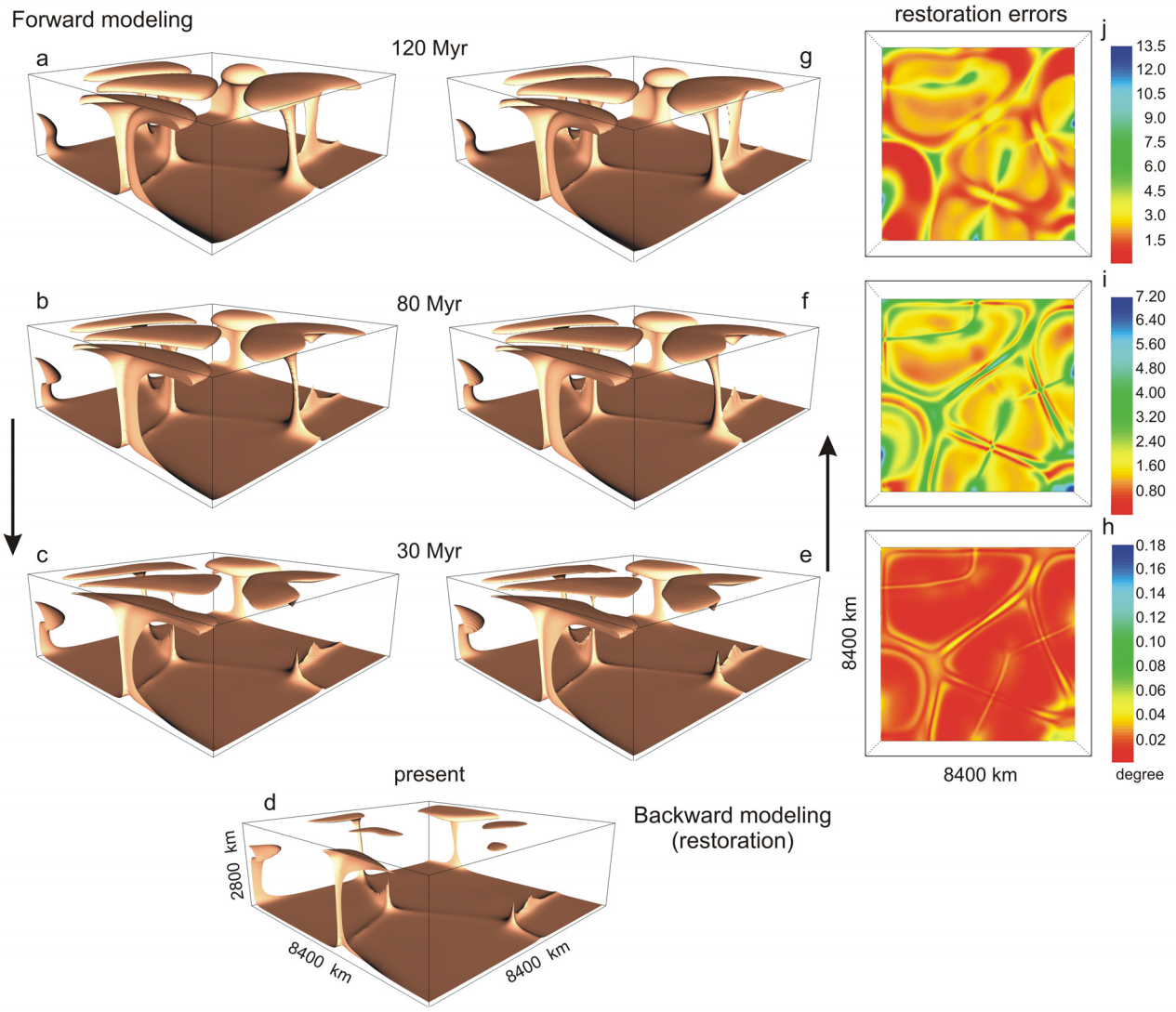

Figure 2. Mantle plume diffusion ($r = 20$ and $Ra = 9.5 \times 10^5$) in the forward modeling at successive diffusion times: (a–d) from 120 Myr ago to the "present" state of the plumes. (e–g) Restored mantle plumes in the backward modeling and (h–j) restoration errors.

[32] To study the effect of thermal diffusion on the restoration of mantle plumes, we develop several independent experiments on mantle plume restoration assigning several Ra values less than the initial Ra by one to 3 orders of magnitude at two values of viscosity ratio r. Figure 3 presents the case of $r = 200$ and $Ra = 9.5 \times 10^3$ and shows several stages in the diffusive decay of the mantle plumes.

[33] The dimensional temperature residuals are within a few degrees for the initial restoration period (Figures 2i and 3h). The computations show that the errors (temperature residuals) get larger the farther the restorations move backward in time (e.g., $\delta T \approx 300$ K at the restoration time of more than 300 Myr, $r = 200$, and $Ra = 9.5 \times 10^3$). Compared to the case of $Ra = 9.5 \times 10^5$, one can see that the residuals become larger as the Rayleigh number decreases or thermal diffusion increases and viscosity ratio increases.

[34] We introduce the critical temperature residual $\delta T_{cr} = 0.2\ \Delta T$ such that the quality of mantle structure recovery is estimated to be bad if $\delta T > \delta T_{cr}$. The quality of the restoration depends on the dimensionless Peclet number $Pe = hu_{\max}\kappa^{-1}$, where $u_{\max}$ is the maximum flow velocity. According to the numerical experiments, the Peclet number corresponding to the critical temperature residual $\delta T_{cr} = 600$ K is $Pe = 10$; Pe should not be less than about 10 for a high-quality plume restoration.

[35] In numerical experiments backward in time we observe an increase in the noise of the restored temperatures with time. *Samarskii et al.* [1997] studied a one-dimensional backward heat diffusion problem and showed that the solution to this problem becomes noisy if the initial temperature guess is slightly perturbed, and the amplitude of this noise increases with the initial perturbations of the temperature guess. They suggest using a special filter to reduce the noise and illustrate the efficiency of the filter. This filter is based on the replacement of iterations (4) by the following iterative scheme:

$$\mathbf{B}(\varphi_{k+1} - \varphi_k) = -\beta_k \nabla J(\varphi_k), \tag{7}$$

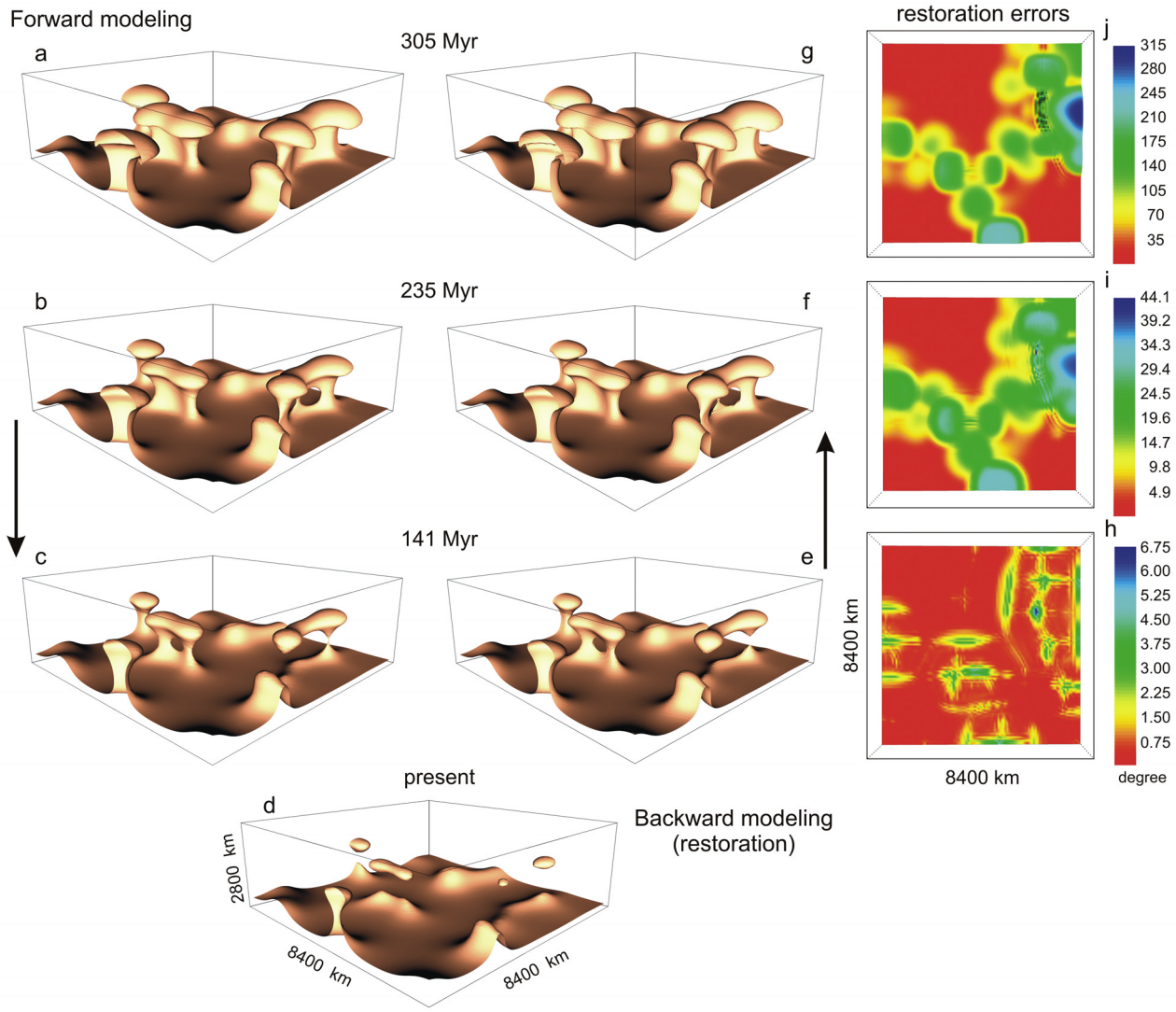

Figure 3. Mantle plume diffusion ($r = 200$ and $Ra = 9.5 \times 10^3$) in the forward modeling at successive diffusion times: (a–d) from 305 Myr ago to the "present" state of the plumes. (e–g) Restored mantle plumes in the backward modeling and (h–j) restoration errors.

where $\mathbf{B}y = y - \nabla^2 y$. Unfortunately, employment of this filter increases the number of iterations to obtain the target temperature and it becomes quite expensive computationally, especially when the model is three-dimensional. Therefore our approach to this problem is to run the model backward to the point of time when the noise becomes relatively large.

6. Performance of the Numerical Algorithm

[36] Here we investigate the impact of diffusion on the performance of our restoration algorithm for various Ra and r. The performance of the algorithm is evaluated in terms of the number of iterations n required to achieve a prescribed relative reduction of $\delta\varphi_n$. Figure 4 presents the evolution of the objective functional $J(\varphi_n)$ and the norm of the gradient of the objective functional $\|\nabla J(\varphi_n)\|$ versus the number of iterations at time about 0.5θ. For other time steps we observe a similar evolution of J and $\|\nabla J\|$.

[37] Both the objective functional and the norm of its gradient show a quite rapid decrease after about 7 iterations for $Ra = 9.5 \times 10^5$ and $r = 20$ (curves 1). The same rapid convergence as a function of adjoint iterations is observed in the *Bunge et al.* [2003] case. As Ra decreases and thermal diffusion increases (curves 2–4) the performance of the algorithm becomes poor: more iterations are needed to achieve the prescribed ε. All curves illustrate that the first 4 to 7 iterations contribute mainly to the reduction of $\delta\varphi_n$. The convergence drops after a relatively small number of iterations. The curves approach the horizontal line with an increase in the number of iterations, because β_k tends to zero with a large number of iterations (see equation (5)). The increase of $\|\nabla J\|$ at $k = 2$ is associated with uncertainty of this gradient at $k = 1$.

[38] Implementation of minimization algorithms requires the evaluation of both the objective functional and its gradient. Each evaluation of the objective functional requires an integration of the model equation (1) with the

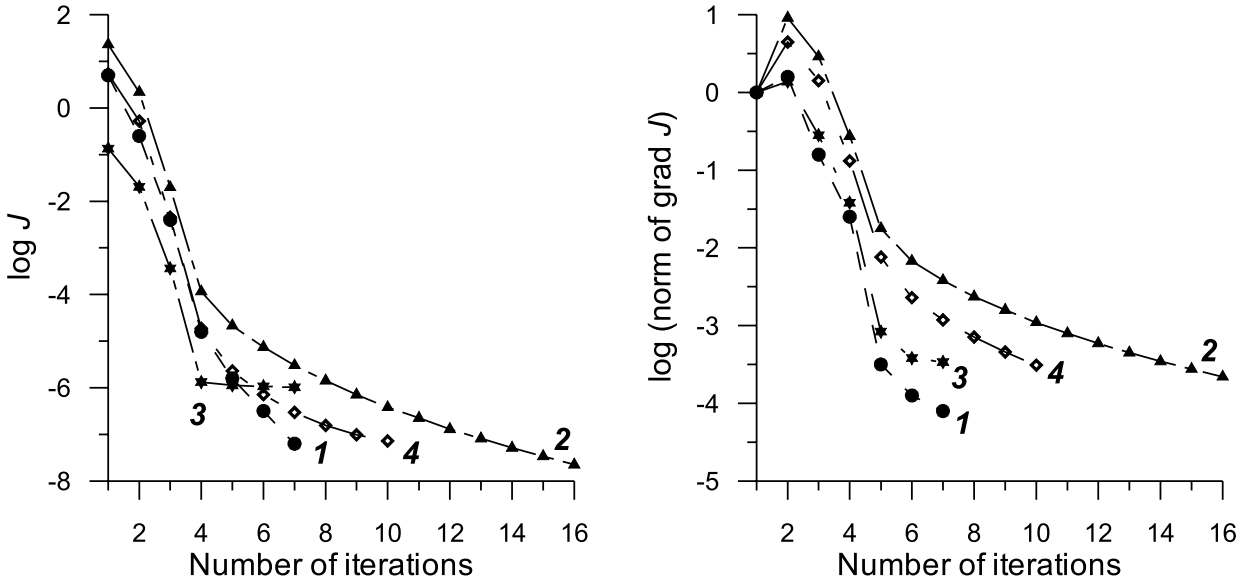

Figure 4. Relative reductions of (left) the objective functional J and (right) the norm of the gradient of J as functions of the number of iterations. Curves indicate 1, $r = 20$, $Ra = 9.5 \times 10^5$; 2, $r = 20$, $Ra = 9.5 \times 10^2$; 3, $r = 200$, $Ra = 9.5 \times 10^3$; and 4, $r = 200$, $Ra = 9.5 \times 10^2$.

appropriate boundary and initial conditions, whereas the gradient is obtained through the backward integration of the adjoint equations (6). The performance analysis shows that the CPU time required to evaluate the gradient J is about the CPU time required to evaluate the objective functional itself, and this is because the direct and adjoint heat problems are described by the same equations.

[39] Despite its simplicity, the minimization algorithm used in this study provides for a rapid convergence and good quality of optimization at high Rayleigh numbers (low thermal diffusion). The convergence rate and the quality of optimization become worse with the decreasing Rayleigh number. The use of the limited-memory quasi-Newton algorithm L-BFGS [*Liu and Nocedal*, 1989] might provide for a better convergence rate and quality of optimization [*Zou et al.*, 1993]. Meanwhile, we note that although an improvement of the convergence rate by using another minimization algorithm (e.g., L-BFGS) will reduce the computational expense associated with the solving of the problem under question, this reduction would be not significant, because the large portion (about 70%) of the computer time is spent to solve the three-dimensional (3-D) Stokes equations.

7. Discussion

7.1. Mantle Plume Heads Yes, Tails No

[40] A plume is hot, narrow mantle upwelling that is invoked to explain hot spot volcanism. In a temperature-dependent viscosity fluid such as the mantle, a plume is characterized by a mushroom-shaped head and a thin tail. Upon impinging under a moving lithosphere, such a mantle upwelling should therefore produce a large amount of melt and successive massive eruption, followed by smaller but long-lived hot spot activity fed from the plume tail [*Morgan*, 1972; *Richards et al.*, 1989; *Sleep*, 1990]. Meanwhile, slowly rising plumes (a buoyancy flux of less than 10^3 kg s^{-1}) coming from the core-mantle boundary should have cooled so much that they would not melt beneath old lithosphere [*Albers and Christensen*, 1996].

[41] A mantle plume is a well-established geological structure in computer modeling and laboratory experiments. Numerical experiments on dynamics of mantle plumes [*Trompert and Hansen*, 1998; *Zhong*, 2005] showed that the number of plumes increases and the rising plumes become thinner with an increase in Rayleigh number. Disconnected thermal plume structures appear in thermal convection at Ra greater than 10^7 [*Hansen et al.*, 1990; *Malevsky et al.*, 1992]. At high Ra (in the hard turbulence regime) thermal plumes are torn off the boundary layer by the large-scale circulation or by nonlinear interactions between plumes [*Malevsky and Yuen*, 1993]. Plume tails can also be disconnected when the plumes are tilted by plate scale flow [e.g., *Olson and Singer*, 1985; *Steinberger and O'Connell*, 1998]. Here we discuss an alternative mechanism for the disconnected mantle plume heads and tails.

[42] Mantle plumes are generated at the top of the thermal boundary layer (TBL), which is produced by conductive heating of the material at the core-mantle boundary (or upper and lower mantle boundary). When the TBL becomes unstable, any perturbation of the TBL top leads to upwelling. Injection of hot material from the source TBL layer into the colder mantle generates strong plumes that are fed for a while from the layer. Colder material overlying the source layer (e.g., portions of lithospheric slabs subducted to the core-mantle boundary) replaces hot material at the locations where the source material is fed into mantle plumes. Some time is required to recover the volume of source material depleted due to plume feeding [*Howard*, 1966]. Because the volume of upwelling material is comparable to the volume of the TBL feeding the mantle plumes, hot material could eventually be exhausted, and mantle plumes would be starved thereafter.

[43] We evaluate the volume V_p of source material that moves into a single plume from the core-mantle boundary

Table 1. Model Parameters and Values

Symbol	Parameter	Value
h	depth of domain, km	2800
g	acceleration due to gravity, m s^{-2}	9.8
P	pressure, Pa	
r	viscosity ratio	20, 200
$t \in [0, \vartheta]$	time, years	
T	temperature, K	
T_{ref}	reference temperature, K	3270
T_{surf}	surface temperature, K	270
T_0	initial dimensionless temperature	
$T*$	initial dimensionless temperature guess	
$\Delta T = T_{\text{ref}} - T_{\text{surf}}$	temperature drop, K	3000
δT	dimensionless temperature residual	
$\mathbf{u} = (u_1, u_2, u_3)$	velocity, cm yr^{-1}	
Pe	Peclet number	
Ra	Rayleigh number	
α	thermal expansivity, K^{-1}	3×10^{-5}
χ	dimensionless temperature at time $t = \vartheta$	
κ	thermal diffusivity, m^2 s^{-1}	10^{-6}
η	viscosity, Pa s	
η_{ref}	reference viscosity, Pa s	8×10^{22}
ρ_{ref}	reference density, kg m^{-3}	4000

over the time interval t_p (required by the plume to reach the base of the lithosphere), and compare that with the volume V_{TBL} of the TBL material conductively generated over the same time interval t_p. For plume height h_p = 2600 km and tail radius r_p ranging from 100 to 200 km, the volume $V_p = \pi r_p^2 h_p$ is estimated to be 0.8 to 3.3 $\times$ 10^8 km^3.

[44] The velocity of plume upwelling

$$w_p = \left(\frac{\alpha g Q}{4\pi c \eta_p}\right)^{1/2}$$

can be estimated analytically from a solution to the boundary layer equations for the steady state flow above a source of heat in a fluid whose viscosity is a temperature-dependent [*Olson et al.*, 1993]. The velocity w_p and the time t_p ($=h_p/w_p$) depend on the plume viscosity η_p as a function of depth and the heat flux $Q = Bc/\alpha$, where B is the buoyancy flux of the plume and c is specific heat. For the typical mantle values given in section 4 (Table 1), B = 3000 to 6000 kg s^{-1}, and $\eta_p = 10^{20}$ to 10^{21} Pa s, the volume $V_{\text{TBL}} = \frac{4}{3}\pi[(r_c + \delta r)^3 - r_c^3]$ (r_c is the radius of the Earth's core, and

$$\delta r = \left(\pi \kappa t_p\right)^{1/2} = \left[\frac{4\pi^3 \kappa^2 h_p^2 \eta}{gB}\right]^{1/4}$$

is the TBL thickness) would range from about 6.3 $\times$ 10^8 to 1.4 $\times$ 10^9 km^3 for the time range of 11 to 48 Myr and the δr range of about 35 to 70 km.

[45] The seismic tomography study [*Montelli et al.*, 2004] has revealed 32 present mantle plumes with radii ranging from 100 to 400 km. Even if only half of the seismically imaged plumes are assumed to have deep mantle roots, we can conclude that the material of the TBL is insufficient to simultaneously feed them. This suggests that only a few mantle plumes can be fed from the TBL at any time and that other plumes are in a phase of thermal diffusive decay. While the discrimination of low-velocity anomalies (seen in seismic tomography models) in active and less active plumes is a challenging problem, laboratory and numerical experiments can provide us with the information.

[46] Recent laboratory experiments on convective instabilities in a layer of fluid with temperature-dependent viscosity and heated from below have shown the generation and evolution of thermal plumes and the transient features of the plumes [*Davaille and Vatteville*, 2005; *Silveira et al.*, 2006]. The temperature difference applied at the lower boundary was chosen such that the Rayleigh number is comparable to that of the Earth's mantle. Initially, a TBL forms at the hot boundary, its thickness increasing by diffusion. When the local Rayleigh number based on the TBL thickness reaches a critical value, the TBL becomes unstable and breaks up to produce plumes [*Howard*, 1966]. A plume reaches the top boundary and spreads laterally. Once the hot TBL has been emptied, the plume tail begins to disappear from the bottom up, leaving only the cooling and shrinking sublithospheric overhangs. The cycle of plume development repeats once the critical thickness of the TBL is reached. The analogue experiments have shown that the mean velocity of the fluid decreases with the maturity of the plumes and hence thermal diffusion becomes a major agent in the heat transfer.

[47] Our numerical results on the diffusive decay of mantle plumes with depleted source regions are in a good agreement with the results of the laboratory experiments. They may have important implications for the interpretation of seismic tomographic images of mantle plumes. Finite frequency seismic tomography images [*Montelli et al.*, 2004] show that a number of plumes extend to midmantle depths but are not visible below these depths. From seismological point of view, the absence of the plume tails could be explained as a combination of several factors [*Romanowicz and Gung*, 2002]: elastic velocities are sensitive to composition as well as temperature; the effect of temperature on velocities decreases with increasing pressure [*Karato*, 1993]; and wavefront healing effects make it difficult to accurately image low-velocity bodies [*Nolet and Dahlen*, 2000]. The "disappearance" of the plume tails can hence be explained as effects of poor tomographic resolution at deeper levels. Apart from this, our results demonstrate the plausibility of finding a great diversity in the morphology of seismically imaged mantle plumes, including plume heads without tails and plumes with tails that are detached from their sources.

[48] The mathematical model of mantle plume dynamics described by a set of equations (1)–(3) is simple, and many complications are omitted. A viscosity increase from the upper to the lower mantle is not included in the model, although it is suggested by studies of the geoid [*Ricard et al.*, 1993], postglacial rebound [*Mitrovica*, 1996], and joint inversion of convection and glacial isostatic adjustment data [*Mitrovica and Forte*, 2004]. The adiabatic heating/cooling term in the heat equation can provide more realistic distribution of temperature in the mantle, especially near the thermal boundary layer. Our model does not include phase transformations [e.g., *Liu et al.*, 1991; *Honda et al.*, 1993a, 1993b; *Harder and Christensen*, 1996], although the phase changes can influence the evolution of mantle plumes retarding/accelerating their ascent. The coefficient of thermal expansion [e.g., *Chopelas and Boehler*, 1989; *Hansen et al.*, 1991, 1993] and the coefficient of thermal conduc-

tivity [e.g., *Hofmeister*, 1999] are not constant in the mantle and vary with depth and temperature. Moreover, if *Badro et al.* [2004] findings of a significant increase in the radiative thermal conductivity at high pressure are relevant to the lower mantle, plume tails should diffuse away even faster than it is predicted by our models.

[49] Mantle plumes exist within the large-scale convective flow, which may disrupt the plumes before they diffuse thermally [e.g., *Richards and Griffiths*, 1988]. *Steinberger* [2000] performed numerical experiments to clarify an interplay between a large-scale mantle flow and mantle plume and hot spot dynamics and showed that during the rise plume tails can be tilted toward large-scale mantle upwellings. Meanwhile, we believe that the possible deformation of plume tails should not significantly alter our results on thermal diffusion of the plumes.

[50] Several reasons constrain us to consider in the present study the simplified mathematical model as the first (principal) step to sophisticated models. The use of the variational data assimilation techniques for the problems of mantle convection began only recently [*Bunge et al.*, 2003; *Ismail-Zadeh et al.*, 2003a]. This technique requires derivation of adjoint equations (to estimate initial temperature conditions in the mantle) each time when the set of the equations is changed. The cost to be paid is in software development since an adjoint model has to be developed. Moreover, since we analyze effects of thermal diffusion on the fate of mantle plumes, we avoid many complications and considered only the most essential component of mantle plume dynamics, namely, temperature-dependent viscosity. While inclusion of these complications and other model refinements are worthwhile, our experiments do show that thermal diffusion plays an important role in the fate of mantle plumes and it provides an explanation for the ''variety'' of mantle plumes observed in seismic tomographic images.

7.2. Assimilation of Present Temperature Derived From Seismic Tomography

[51] The variational assimilation of synthetic data (mantle plumes generated by computer simulations) showed a possibility to restore strong features of the plumes after their thermal diffusion. In this section we illustrate how real (no synthetic) present crust/mantle temperature can be assimilated into the geological past. For this aim we use recent teleseismic body wave tomography data, which image the lithosphere and asthenosphere for the southeastern Carpathians [*Martin et al.*, 2005]. We should note that the region is not associated with a mantle plume activity and chosen because of high-resolution seismic tomography data made available to the authors.

[52] The seismic tomographic model of the region consists of eight layers of different thickness (from 15 to 50 km), which are each subdivided laterally into 42×42 km^2 blocks [*Martin et al.*, 2005]. To restrict numerical errors in our data assimilation we smooth the velocity anomaly data using spline interpolations between the blocks and the layers. To convert the *P wave* seismic velocity anomalies beneath the region into temperature we model initially synthetic *P wave* seismic velocities considering the effects of anharmonicity (composition), anelasticity and partial melting on the seismic velocities [*Ismail-Zadeh et al.*, 2005]. The anharmonic (frequency-independent and nonattenuating) part of the synthetic velocities is calculated on the basis of published data on laboratory measurements of density and elastic parameters of the main rock-forming minerals [*Bass*, 1995] at various thermodynamic conditions for the composition of the crust and mantle (57.9% Ol, 16.3% CPx, 13.5% Opx, and 12.3% Gt [*Green and Falloon*, 1998]) and the slab (69% Ol, 10% CPx, 19% Opx, and 2% Gt [*Agee*, 1993]). Once the synthetic velocities are calculated for a first-guess temperature, an iteration process is used to find the ''true'' temperature, minimizing the difference between the synthetic and ''observed'' (in seismic tomography experiments) velocities. The temperature in the shallow levels of the region is constrained from measured surface heat flux corrected for paleoclimate changes and for the effects of sedimentation [*Demetrescu et al.*, 2001]. Figure 5a illustrates several depth slices of the present temperature model derived from the seismic tomography data.

[53] We assimilate the present temperature data into the geological past to restore the prominent thermal features of the Earth's structures in the region. We use the following parameters in this case study: $h = 670$ km, the aspect ratio (ratio between horizontal and vertical lengths of the model) is 1.5, $r = 1000$, $\Delta T = 1700$ K, $\rho_{ref} = 3400$ kg m^{-3}, $\eta_{ref} = 10^{21}$ Pa s, $Ra = 5.2 \times 10^5$. Other parameters are the same (see Table 1). The equations and boundary conditions are defined in sections 2 and 3. To reduce the numerical noise in the data assimilation, we regularize the solution by using the quasi-reversibility method by *Lattes and Lions* [1969]. Figure 5b shows the temperature restored to 22 Myr ago.

[54] Early Miocene subduction beneath the Carpathian arc and subsequent gentle continental collision transported cold and dense lithospheric material into the hotter mantle [*Sperner and The CRC 461 Team*, 2005]. The cold (blue) region seen at the 20 km slice of the restored temperature (Figure 5b) can be interpreted as a crustal portion of a lithospheric slab. The structure is almost invisible at the relevant slice of the present temperature, because the slowly descending cold slab has been warmed up (and hence has faded away) due to thermal diffusion since an active slab subduction in the region has ended about 10 Myr ago [*Csontos et al.*, 1992]. Thermal conduction in the shallow Earth (where viscosity is high) plays a significant part in heat transfer compared to thermal convection. The deeper we look into the region (see the slices at depths of 60 km and 130 km in Figure 5b), the larger are effects of thermal advection compared to diffusion: the cold (dense) lithosphere has moved upward to the place where it has been in the Miocene times. At 280 km depth a shape of the colder slab is clearly visible at the slice of the present temperature (Figure 5a) and practically invisible at the slice of the restored temperature (Figure 5b), because the slab did not reach the depth 22 Myr ago.

[55] Thus the assimilation of the present temperature derived from seismic tomography data shows that prominent thermal mantle structures can be restored from their present diffused stage.

8. Conclusion

[56] In this paper, models of mantle plume weakening due to thermal diffusion have been analyzed. Injection of hot material from the thermal source layer into the colder

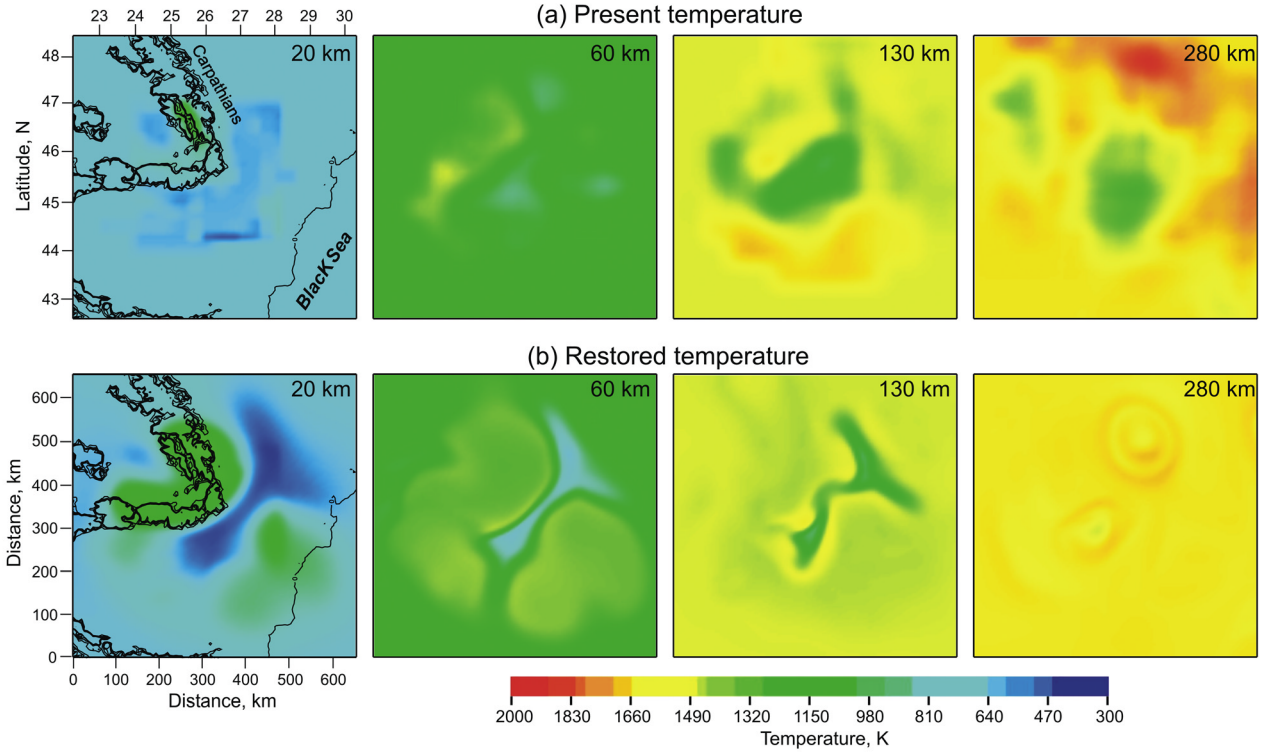

Figure 5. Present and restored (to 22 Myr ago) temperature beneath the southeastern Carpathians at depths of 20, 60, 130, and 280 km. (a) Temperatures derived from *P* wave velocity anomalies. (b) Temperature restored by data assimilation. Isolines present the surface topography.

mantle generates strong plumes that are fed from the source layer for a while. However, the feeding from the source layer can weaken with time and then thermal diffusion takes over and controls the subsequent evolution of the mantle plumes. The plumes begin to diffuse away and the plume tails are the first structures to disappear. The tails of different plumes vanish at different times depending on the geometry of the tails. The morphological diversity of the plumes predicted by the numerical experiments is similar to the plume diversity observed in seismic tomographic images [*Montelli et al.*, 2004; *Zhao*, 2004].

[57] We have also studied how the restoration process (data assimilation algorithm) works in recovering strong features of mantle plumes after they have weakened by thermal diffusion and in the presence of a large depth gradient of mantle viscosity. The restoration process becomes poor as both diffusion and viscosity gradient increase. For a given range of Rayleigh number and two values of the viscosity gradient, the convergence rate of the objective functional shows a large variation, which implies that the performance is very sensitive to the magnitude of both diffusion and viscosity gradient.

[58] The present temperature obtained from high-resolution teleseismic tomography data for the southeastern Carpathians has been assimilated into the geological past. Results of this case study suggest that the data assimilation can be used to restore initial mantle temperatures and can allow revealing prominent thermal structures in the mantle from their present diffused stage. A part of the geophysical community may maintain skepticism about the assimilation of present mantle-related data to the geological past. This skepticism may partly have its roots in our poor knowledge of the Earth's present structure and its physical properties, which cannot allow for rigorous numerical paleoreconstructions of the mantle evolution. An increase in the accuracy of seismic tomography inversions and geodetic measurements, improvements in the knowledge of gravity and geothermal fields, and more complete experimental data on the physical and chemical properties of mantle rocks will facilitate mantle reconstructions.

Appendix A: Challenges in Variational Data Assimilation for Thermoconvective Flow in the Mantle

[59] Although the variational data assimilation technique described above can theoretically be applied to many problems in mantle and lithosphere dynamics, a practical implementation of the technique for modeling of real geodynamic processes backward in time (to restore the temperature and flow pattern in the past) is not a simple task. Smoothness of the initial data (present temperature) and of the target temperature (restored temperature in the past) is an important factor in backward modeling. Moreover, a choice of the initial temperature guess φ_0 in iteration scheme (4) is not trivial.

A1. On the Smoothness of the Initial Temperature

[60] The solution $T(\vartheta, \cdot\,; \varphi)$ of the heat problem (1) is a sufficiently smooth function and belongs to space $L_2(\Omega)$. The present temperature χ_δ derived from the seismic tomography is a representation of the exact temperature χ of the Earth and so it must also belong to this space and

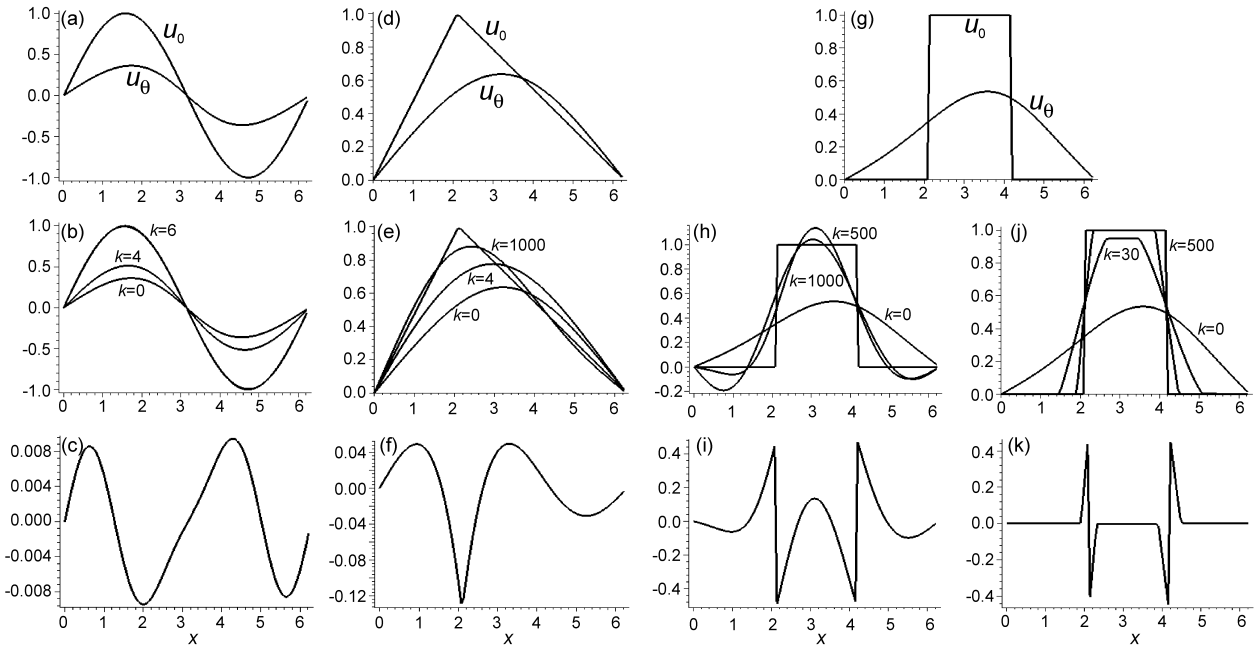

Figure A1. Recovering function u_0 from the smooth guess function u_θ. (a–c) The sufficiently smooth u_0; (d–f) continuous piecewise smooth function u_0; and (g–k) discontinuous function u_0. Plots of u_0 and u_θ are presented in Figures A1a, A1d, and A1g; successive approximations to u_0 at Figures A1b, Figures A1e, Figures A1h, and Figures A1j; and the residual functions in Figures A1c, Figures A1f, Figures A1i, and Figures A1k.

hence be rather smooth; otherwise, the objective functional J cannot be defined. Therefore before any assimilation of the present temperature data can be attempted, the data must be smoothed. The smoothing of the present temperature improves the convergence of the iterations. However, there are still some numerical issues associated with the solution of the improperly posed problem (we remind the reader that the inverse problem of thermal convection is improperly posed [e.g., *Tikhonov and Arsenin*, 1977]).

[61] If the initial temperature guess φ_0 is a smooth function, all successive temperature iterations φ_k in scheme (4) should be smooth functions too, because the gradient of the objective functional ∇J is a smooth function since it is the solution to the adjoint problem (6). The temperature iterations φ_k are disturbed by small computational errors, which are inherent in any numerical experiment (see Appendix B). These perturbations grow with time unless the iteration scheme (7) or a similar one [*Samarskii and Vabischevich*, 1995] is used as discussed in section 5. Another possibility is to use the quasi-reversibility method [*Lattes and Lions*, 1969] to regularize a temperature field or high-order adjoint techniques [*Alekseev and Navon*, 2001].

[62] A choice of the initial temperature guess φ_0 (smooth versus discontinuous functions) influences the convergence of the iterations. There are however no general "recipes" for the choice of the initial temperature guess, and this depends mainly on the experience of computer modelers in solving such numerical problems.

A2. On the Smoothness of the Target Temperature

[63] If mantle temperature in the geological past was not a smooth function of space variables, recovery of this temperature using the technique described in this paper is not effective because the iterations converge very slowly to the target temperature. Here we explain the problem of recovering the initial mantle temperature at the time of plume onset on the basis of three one-dimensional model tasks: restoration of a smooth, piecewise smooth and discontinuous target function. We note that the temperature in the Earth's mantle is not a discontinuous function but its shape can be close to a step function.

[64] We consider that the dynamics of a physical system is described by the Burgers equation $u_t + uu_x = u_{xx}$, $0 \le t \le 1$, $0 \le x \le 2\pi$ with the boundary conditions $u(t, 0) = 0$, $u(t, 2\pi) = 0$, $0 \le t \le 1$ and the condition $u_\theta = u(1, x; u_0)$, $0 \le x \le 2\pi$ at $t = 1$, where the variable u can denote temperature. The problem is to recover the function $u_0 = u_0(x)$, $0 \le x \le 2\pi$ at $t = 0$ (the state in the past) from the function $u_\theta = u_\theta(x)$, $0 \le x \le 2\pi$ at $t = 1$ (its present state). The finite difference approximations and the variational method are applied to the Burgers equation with the appropriate boundary and initial conditions.

A2.1. Task 1

[65] Consider the sufficiently smooth function $u_0 = \sin(x)$, $0 \le x \le 2\pi$. The functions u_0 and u_θ are shown in Figure A1a. Figures A1b and A1c illustrate the iterations φ_k using the iterative scheme similar to (4) for $k = 0$, 4, 6 and the residual $r_6(x) = u_0(x) - \varphi_6(x)$, $0 \le x \le 2\pi$, respectively. We see that iterations converge rather rapid for the sufficiently smooth target function.

A2.2. Task 2

[66] Now consider the continuous piecewise smooth function $u_0 = 3x/(2\pi)$, $0 \le x \le 2\pi/3$ and $u_0 = 3/2 - 3x/(2\pi)$, $2\pi/3 \le x \le 2\pi$. Figure A1 presents the functions u_0 and

u_0 (Figure A1d), the successive approximations φ_k for $k = 0$, 4, 1000 (Figure A1e), and the residual $r_{1000}(x) = u_0 x) - \varphi_{1000}(x)$, $0 \le x \le 2\pi$ (Figure A1f), respectively. This example shows that a large number of iterations is required to reach the target function.

A2.3. Task 3

[67] Consider the discontinuous function u_0, which takes 1 at $2\pi/3 \le x \le 4\pi/3$ and 0 in other points of the closed interval $0 \le x \le 2\pi$. Figure A1 presents the functions u_0 and u_0 (Figure A1g), the successive approximations φ_k for $k = 0$, 500, 1000 (Figure A1h), and the residual $r_{1000}(x) = u_0(x) - \varphi_{1000}(x)$, $0 \le x \le 2\pi$ (Figure A1e), respectively. We see that convergence to the target temperature is very poor.

[68] To improve the convergence to the target function, a modification of the variational method based on a priori information about a desired solution was suggested by *Korotkii and Tsepelev* [2003]. Figure A1j shows the successive approximations $\widetilde{\varphi}_k$ for $k = 0$, 30, 500, and Figure A1k shows the residual $\widetilde{r}_{500}(x) = u_0(x) - \widetilde{\varphi}_{500}(x)$, $0 \le x \le 2\pi$, respectively. The approximations $\widetilde{\varphi}_k$ based on the method of gradient projection [*Vasiliev*, 2002] converge to the target solution better than approximations generated by equation (4).

Appendix B: Errors in Forward and Backward Modeling

[69] A numerical model has three kinds of variables: state variables, input variables, and parameters. State variables describe the physical properties of the medium (velocity, pressure, temperature) and depend on time and space. Input variables have to be provided to the model (initial or boundary conditions), most of the time these variables are not directly measured but they can be estimated through data assimilation. Most models contain also a set of parameters (e.g., viscosity, thermal diffusivity), which have to be tuned to adjust the model to the observations. All the variables can be polluted by errors.

[70] There are three kinds of systematic errors in numerical modeling of geodynamical problems: model, discretization, and iteration errors. Model errors are associated with the idealization of Earth dynamics by a set of conservation equations governing the dynamics. The model errors are defined as the difference between the actual Earth dynamics and the exact solution of the mathematical model. Discretization errors are defined as the difference between the exact solution of the conservation equations and the exact solution of the algebraic system of equations obtained by discretizing these equations. Also, iteration errors are defined as the difference between the iterative and exact solutions of the algebraic system of equations. It is important to be aware of the existence of these errors, and even more to try to distinguish one from another.

[71] Apart from the errors associated with the numerical modeling, another two components of errors are essential when mantle temperature data are assimilated into the past: (1) data misfit associated with the uncertainties in the present temperature distribution in the Earth's mantle and (2) errors associated with the uncertainties in initial and boundary conditions. Since there are no direct measurements of mantle temperatures, the temperatures can be estimated indirectly either from seismic wave (and their anomalies), geochemical analysis or through the extrapolation of surface heat flow observations. Many models of mantle temperature are based on the conversion of seismic tomography data into temperature. Meanwhile, a seismic tomography image of the Earth's mantle is a model indeed and incorporates its own model errors. Another source of uncertainty comes from the choice of mantle compositions in the modeling of mantle temperature from the seismic velocities. Therefore, if the present mantle temperature models are biased, information on temperature can be improperly propagated to the geological past.

[72] The temperature at the lower boundary of the model domain we used in forward and backward numerical modeling is, of course, an approximation to the real temperature, which is unknown and may change over time at this boundary. Hence errors associated with the knowledge of the temperature (or heat flux) evolution at the core-mantle boundary are another essential component of errors, which can be propagated into the past during the data assimilation.

[73] In numerical modeling sensitivity analysis assists in understanding the stability of the model solution to small perturbations in input variables or parameters. For instance, if we consider mantle temperature in the past as a solution to the backward model, what will be its variation if there is some perturbation on the inputs of the model (e.g., present temperature data)? The gradient of the objective functional with respect to input parameters in variational data assimilation gives the first-order sensitivity coefficients. The second-order adjoint sensitivity analysis presents some challenge associated with cumbersome computations of the product of the Hessian matrix of the objective functional with some vector [*Le Dimet et al.*, 2002], and hence it is omitted in our study. *Hier-Majumder et al.* [2006] performed the first-order sensitivity analysis for two-dimensional problems of thermoconvective flow in the mantle. See *Cacuci* [2003] and *Cacuci et al.* [2005] for more detail on sensitivity and uncertainty analysis.

[74] **Acknowledgments.** We thank A. Davaille, M. Ghil, S. Honda, N. Ribe, H. Wilhelm, and T. Yanovskaya for discussions on thermal plume diffusion, data assimilation, and on geothermal and seismic tomography modeling. We are very grateful to H.-P. Bunge, I. M. Navon, anonymous reviewers and the Associate Editor, who provided careful reviews that significantly improved an initial version of the manuscript. The research was supported by the German Research Council (DFG-Wi-687/18-1), Russian Foundation of Basic Research (RFBR-05-01-00098), and the Russian Academy of Sciences (OH3-260603-879) grants.

References

Agee, C. B. (1993), Petrology of the mantle transition zone, *Annu Rev. Earth Planet. Sci.*, *21*, 19–41.

Albers, M., and U. R. Christensen (1996), The excess temperature of plumes rising from the core-mantle boundary, *Geophys. Res. Lett.*, *23*, 3567–3570.

Alekseev, A. K., and I. M. Navon (2001), The analysis of an ill-posed problem using multiscale resolution and second order adjoint techniques, *Comput.Methods Appl. Mech. Eng.*, *190*, 1937–1953.

Badro, J., J.-P. Rueff, G. Vanko, G. Monaco, G. Figuet, and F. Guyot (2004), Electronic transitions in perovskite: Possible nonconvecting layers in the lower mantle, *Science*, *305*, 383–386.

Bass, J. D. (1995), Elasticity of minerals, glasses, and melts, in *Mineral Physics and Crystallography, A Handbook of Physical Constants*, *AGU Ref. Shelf*, vol. 2, edited by T. J. Ahrens, pp. 45–63, AGU, Washington, D. C.

Bennett, A. F. (1992), *Inverse Methods in Physical Oceanography*, 346 pp., Cambridge Univ. Press, New York.

Boussinesq, J. (1903), *Theorie Analytique de la Chaleur*, vol. 2, 172 pp., Elsevier, New York.

Bunge, H.-P., M. A. Richards, C. Lithgow-Bertelloni, J. R. Baumgardner, S. P. Grand, and B. Romanowicz (1998), Time scales and heterogeneous structure in geodynamic Earth models, *Science*, *280*, 91–95.

Bunge, H.-P., M. A. Richards, and J. R. Baumgardner (2002), Mantle circulation models with sequential data-assimilation: Inferring present-day mantle structure from plate motion histories, *Philos. Trans. R. Soc. London, Ser. A*, *360*, 2545–2567.

Bunge, H.-P., C. R. Hagelberg, and B. J. Travis (2003), Mantle circulation models with variational data assimilation: Inferring past mantle flow and structure from plate motion histories and seismic tomography, *Geophys. J. Int.*, *152*, 280–301.

Busse, F. H., et al. (1993), 3D convection at infinite Prandtl number in Cartesian geometry–a benchmark comparison, *Geophys. Astrophys. Fluid Dyn.*, *75*, 39–59.

Cacuci, D. G. (2003), *Sensitivity and Uncertainty Analysis*, vol. I, *Theory*, 285 pp., CRC, Boca Raton, Fla.

Cacuci, D. G., M. Ionescu-Bujor, and I. M. Navon (2005), *Sensitivity and Uncertainty Analysis. Volume II: Applications to Large-Scale Systems*, 368 pp., CRC, Boca Raton, Fla.

Chandrasekhar, S. (1961), *Hydrodynamic and Hydromagnetic Stability*, 654 pp., Oxford Univ. Press, New York.

Chopelas, A., and R. Boehler (1989), Thermal expansion measurements at very high pressure, systematics and a case for a chemically homogeneous mantle, *Geophys. Res. Lett.*, *16*, 1347–1350.

Condie, K. C. (2001), *Mantle Plumes and Their Record in Earth History*, 306 p., Cambridge Univ. Press, New York.

Csontos, L., A. Nagymarosy, F. Horvath, and M. Kovac (1992), Tertiary evolution of the intra-Carpathian area: A model, *Tectonophysics*, *208*, 221–241.

Davaille, A. (1999), Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle, *Nature*, *402*, 756–760.

Davaille, A., and J. Vatteville (2005), On the transient nature of mantle plumes, *Geophys. Res. Lett.*, *32*, L14309, doi:10.1029/2005GL023029.

Demetrescu, C., S. B. Nielsen, M. Ene, D. Z. Serban, G. Polonic, M. Andreescu, A. Pop, and N. Balling (2001), Lithosphere thermal structure and evolution of the Transylvanian Depression: Insight from new geothermal measurements and modelling results, *Phys. Earth Planet. Inter.*, *126*, 249–267.

Fletcher, R., and C. M. Reeves (1964), Function minimization by conjugate gradients, *Comput. J.*, *7*, 149–154.

Forte, A. M., and J. X. Mitrovica (2001), Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data, *Nature*, *410*, 1049–1056.

Ghil, M., and P. Malanotte-Rizzoli (1991), Data assimilation in meteorology and oceanography, *Adv. Geophys.*, *33*, 141–266.

Green, D. H., and T. J. Falloon (1998), Pyrolite: A Ringwood concept and its current expression, in *The Earth's Mantle*, edited by I. Jackson, pp. 311–378, Cambridge Univ. Press, New York.

Hansen, U., D. A. Yuen, and S. E. Kroening (1990), Transition to hard turbulence in thermal convection at infinite Prandtl number, *Phys. Fluids A*, *2*(12), 2157–2163.

Hansen, U., D. A. Yuen, and S. E. Kroening (1991), Effects of depth-dependent thermal expansivity on mantle circulations and lateral thermal anomalies, *Geophys. Res. Lett.*, *18*, 1261–1264.

Hansen, U., D. A. Yuen, S. E. Kroening, and T. B. Larsen (1993), Dynamical consequences of depth-dependent thermal expansivity and viscosity on mantle circulations and thermal structure, *Phys. Earth Planet. Inter.*, *77*, 205–223.

Harder, H., and U. R. Christensen (1996), A one-plume model of Martian mantle convection, *Nature*, *380*, 507–509.

Hier-Majumder, C. A., B. J. Travis, E. Belanger, G. Richard, A. P. Vincent, and D. A. Yuen (2006), Efficient sensitivity analysis for flow and transport in the Earth's crust and mantle, *Geophys. J. Int.*, in press.

Hofmeister, A. M. (1999), Mantle values of thermal conductivity and the geotherm from photon lifetimes, *Science*, *283*, 1699–1706.

Honda, S., S. Balachandar, D. A. Yuen, and D. Reuteler (1993a), Three-dimensional mantle dynamics with an endothermic phase transition, *Geophys. Res. Lett.*, *20*, 221–224.

Honda, S., D. A. Yuen, S. Balachandar, and D. Reuteler (1993b), Three-dimensional instabilities of mantle convection with multiple phase transitions, *Science*, *259*, 1308–1311.

Howard, L. N. (1966), Convection at high Rayleigh number, in *Applied Mechanics, Proc. of the 11th Intl Congress of Applied Mechanics, Munich, Germany 1964*, edited by H. Goertler and P. Sorger, pp. 1109–1115, Springer, New York.

Ismail-Zadeh, A. T., A. I. Korotkii, B. M. Naimark, and I. A. Tsepelev (2001), Numerical simulation of three-dimensional viscous flows with gravitational and thermal effects, *Comput. Math. Math. Phys.*, *41*, 1399–1415.

Ismail-Zadeh, A. T., A. I. Korotkii, B. M. Naimark, and I. A. Tsepelev (2003a), Three-dimensional numerical simulation of the inverse problem of thermal convection, *Comput. Math. Math. Phys.*, *43*, 587–599.

Ismail-Zadeh, A. T., A. I. Korotkii, and I. A. Tsepelev (2003b), Numerical approach to solving problems of slow viscous flow backwards in time, in *Computational Fluid and Solid Mechanics*, edited by K. J. Bathe, pp. 938–941, Elsevier, New York.

Ismail-Zadeh, A., G. Schubert, I. Tsepelev, and A. Korotkii (2004), Inverse problem of thermal convection: Numerical approach and application to mantle plume restoration, *Phys. Earth Planet. Inter.*, *145*, 99–114.

Ismail-Zadeh, A., B. Mueller, and G. Schubert (2005), Three-dimensional modeling of present-day tectonic stress beneath the earthquake-prone southeastern Carpathians based on integrated analysis of seismic, heat flow, and gravity observations, *Phys. Earth Planet. Inter.*, *149*, 81–98.

Jellinek, M. A., and M. Manga (2002), The influence of a chemical boundary layer on the fixity, spacing and lifetime of mantle plumes, *Nature*, *418*, 760–763.

Kalnay, E. (2003), *Atmospheric Modeling, Data Assimilation and Predictability*, 341 pp., Cambridge Univ. Press, New York.

Karato, S.-I. (1993), Importance of anelasticity in the interpretation of seismic tomography, *Geophys. Res. Lett.*, *20*, 1623–1626.

Korotkii, A. I., and I. A. Tsepelev (2003), Solution of a retrospective inverse problem for one nonlinear evolutionary model, in *Proceedings of the Steklov Institute of Mathematics, Suppl. 2*, pp. 80–94, Nauka, Moscow.

Lattes, R., and J. L. Lions (1969), *The Method of Quasi-Reversibility: Applications to Partial Differential Equations*, 388 pp., Elsevier, New York.

Le Dimet, F.-X., I. M. Navon, and D. N. Daescu (2002), Second-order information in data assimilation, *Mon. Weather Rev.*, *130*, 629–648.

Leitch, A. M., V. Steinbach, and D. A. Yuen (1996), Centerline temperature of mantle plumes in various geometries: Incompressible flow, *J. Geophys. Res.*, *101*, 21,829–21,846.

Liu, D. C., and J. Nocedal (1989), On the limited memory BFGS method for large scale optimization, *Math. Program.*, *45*, 503–528.

Liu, M., D. A. Yuen, W. Zhao, and S. Honda (1991), Development of diapiric structures in the upper mantle due to phase transitions, *Science*, *252*, 1836–1839.

Malevsky, A. V., and D. A. Yuen (1993), Plume structures in the hard-turbulent regime of three-dimensional infinite Prandtl number convection, *Geophys. Res. Lett.*, *20*, 383–386.

Malevsky, A. V., D. A. Yuen, and L. M. Weyer (1992), Viscosity and thermal fields associated with strongly chaotic non-Newtonian thermal convection, *Geophys. Res. Lett.*, *19*, 127–130.

Martin, M., J. R. R. Ritter, and the CALIXTO Working Group (2005), High-resolution teleseismic body-wave tomography beneath SE Romania–I. Implications for three-dimensional versus one-dimensional crustal correction strategies with a new crustal velocity model, *Geophys. J. Int.*, *162*, 448–460.

McDonald, A. (1984), Accuracy of multi-upstream semi-Lagrangian advective schemes, *Mon. Weather Rev.*, *112*, 1267–1279.

McLaughlin, D. (2002), An integrated approach to hydrologic data assimilation: Interpolation, smoothing, and forecasting, *Adv. Water Resour.*, *25*, 1275–1286.

Mitrovica, J. X. (1996), Haskell (1935) revisited, *J. Geophys. Res.*, *101*, 555–569.

Mitrovica, J. X., and A. M. Forte (2004), A new inference of mantle viscosity based upon joint inversion of convection and glacial isostatic adjustment data, *Earth Planet. Sci. Lett.*, *225*, 177–189.

Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung (2004), Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, *303*, 338–343.

Moore, W. B., G. Schubert, and P. Tackley (1998), Three-dimensional simulations of plume–lithosphere interaction at the Hawaiian Swell, *Science*, *279*, 1008–1011.

Morgan, W. J. (1972), Plate motions and deep convection, *Geol. Soc. Am. Mem.*, *132*, 7–22.

Navon, I. M., X. Zou, J. Derber, and J. Sela (1992), Variational data assimilation with an adiabatic version of the NMC spectral model, *Mon. Weather Rev.*, *120*, 1433–1446.

Nolet, G., and F. A. Dahlen (2000), Wave front healing and the evolution of seismic delay times, *J. Geophys. Res.*, *105*, 19,043–19,054.

Olson, P., and H. Singer (1985), Creeping plumes, *J. Fluid Mech*, *158*, 511–531.

Olson, P., G. Schubert, and C. Anderson (1993), Structure of axisymmetric mantle plumes, *J. Geophys. Res.*, *98*, 6829–6844.

Ribe, N. M., and U. Christensen (1994), Three-dimensional modeling of plume–lithosphere interaction, *J. Geophys. Res.*, *99*, 669–682.

Ricard, Y., M. A. Richards, C. Lithgow-Bertelloni, and Y. Le Stunff (1993), A geodynamic model of mantle density heterogeneity, *J. Geophys. Res.*, *98*, 21,895–21,909.

Richards, M. A., and R. W. Griffiths (1988), Deflection of plumes by mantle shear flow: Experimental results and a simple theory, *Geophys. J. R. Astron. Soc.*, *94*, 367–376.

Richards, M. A., R. A. Duncan, and V. Courtillot (1989), Flood basalts and hot spot tracks: Plume heads and tails, *Science*, *246*, 103–107.

Ritsema, J., H. J. van Heijst, and J. H. Woodhouse (1999), Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, *286*, 1925–1928.

Romanowicz, B., and Y. Gung (2002), Superplumes from the core-mantle boundary to the lithosphere: Implications for heat flux, *Science*, *296*, 513–516.

Samarskii, A. A., and P. N. Vabishchevich (1995), *Computational Heat Transfer*, vol. 1, *Mathematical Modelling*, 370 pp., John Wiley, Hoboken, N. J.

Samarskii, A. A., and P. N. Vabishchevich (2004), *Numerical Methods for Solving Inverse Problems of Mathematical Physics*, 478 pp., URSS, Moscow.

Samarskii, A. A., P. N. Vabishchevich, and V. I. Vasil'ev (1997), Iterative solution of a retrospective inverse problem of heat conduction, *Math. Model.*, *9*(5), 119–127.

Schubert, G., D. L. Turcotte, and P. Olson (2001), *Mantle Convection in the Earth and Planets*, 940 p., Cambridge Univ. Press, New York.

Silveira, G., E. Stutzmann, A. Davaille, J.-P. Montagner, L. Mendes-Victor, and A. Sebai (2006), Azores hotspot signature in the upper mantle, *J. Volcanol. Geotherm. Res.*, in press.

Sleep, N. H. (1990), Hotspots and mantle plumes: Some phenomenology, *J. Geophys. Res.*, *95*, 6715–6736.

Sperner, B., and The CRC 461 Team (2005), Monitoring of slab detachment in the Carpathians, in *Perspectives in Modern Seismology*, *Lect. Notes Earth Sci.*, vol. 105, edited by F. Wenzel, pp. 187–202, Springer, New York.

Steinberger, B. (2000), Plumes in a convecting mantle: Models and observations for individual hotspots, *J. Geophys. Res.*, *105*, 11,127–11,152.

Steinberger, B., and R. J. O'Connell (1998), Advection of plumes in mantle flow: Implications for hotspot motion, mantle viscosity and plume distribution, *Geophys. J. Int.*, *132*, 412–434.

Talagrand, O. (1997), Assimilation of observations: An introduction, *J. Meteorol. Soc. Jpn.*, *75*, 191–209.

Tikhonov, A. N., and V. Y. Arsenin (1977), *Solution of Ill-Posed Problems*, 258 pp., Winston, Washington, D. C.

Trompert, R. A., and U. Hansen (1998), On the Rayleigh number dependence of convection with a strongly temperature-dependent viscosity, *Phys. Fluids*, *10*, 351–360.

Vasiliev, F. P. (2002), *Methods of Optimization* (in Russian), 824 pp., Factorial Press, Moscow.

Wunsch, C. (1996), *The Ocean Circulation Inverse Problem*, Cambridge Univ. Press, 442 pp., New York.

Zhao, D. (2004), Global tomographic images of mantle plumes and subducting slabs: Insight into deep Earth dynamics, *Phys. Earth Planet. Inter.*, *146*, 3–34.

Zhong, S. (2005), Dynamics of thermal plumes in three-dimensional isoviscous thermal convection, *Geophys. J. Int.*, *162*, 289–300.

Zou, X., I. M. Navon, M. Berger, K. H. Phua, T. Schlick, and F. X. Le Dimet (1993), Numerical experience with limited-memory quasi-Newton and truncated Newton methods, *SIAM J. Optimization*, *3*(3), 582–608.

A. Ismail-Zadeh, Geophysikalishes Institut, Universität Karlsruhe, Hertzstr. 16, Karlsruhe D-76187, Germany. (alik.ismail-zadeh@gpi.uka.de)

A. Korotkii and I. Tsepelev, Institute of Mathematics and Mechanics, Ural Branch, Russian Academy of Sciences, S. Kovalevskoy ul. 16, Yekaterinburg 620219, Russia.

G. Schubert, Department of Earth and Space Sciences, University of California, 3806 Geology Building, 595 Charles Young Drive East, Los Angeles, CA 90095-1567, USA.